

Wolf, Willi

Beurteilungsmodelle bei Klassenarbeiten. Bestandsaufnahme schulischer Praxis in der Sekundarstufe I

Zeitschrift für Pädagogik 30 (1984) 1, S. 115-130



Quellenangabe/ Reference:

Wolf, Willi: Beurteilungsmodelle bei Klassenarbeiten. Bestandsaufnahme schulischer Praxis in der Sekundarstufe I - In: Zeitschrift für Pädagogik 30 (1984) 1, S. 115-130 - URN: urn:nbn:de:0111-pedocs-142927 - DOI: 10.25656/01:14292

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-142927>

<https://doi.org/10.25656/01:14292>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Zeitschrift für Pädagogik

Jahrgang 30 – Heft 1 – Februar 1984

I. Essay

KONRAD WÜNSCHE

Ludwigs Einschulung.
Eine pädagogische Unterhaltung 1

II. Thema: Allgemeine Pädagogik

EGON SCHÜTZ

Einige Überlegungen zur Fragwürdigkeit systematischer
Pädagogik 17

ROLF HUSCHKE-RHEIN

Über die Zukunft der Allgemeinen Pädagogik. Systema-
tische und systemökologische Überlegungen 31

HEINZ-ELMAR TENORTH

Berufsethik, Kategorialanalyse, Methodenreflexion.
Zum historischen Wandel des „Allgemeinen“ in der
wissenschaftlichen Pädagogik 49

III. Thema: Schriftspracherwerb

HANS BRÜGELMANN

Lesen- und Schreibenlernen als Denkentwicklung. Vor-
aussetzungen eines erfolgreichen Schrifterwerbs 69

MECHTHILD DEHN

Lernschwierigkeiten beim Schriftspracherwerb. Krite-
rien zur Analyse des Leselernprozesses und zur Diffe-
renzierung von Lernschwierigkeiten 93

IV. Diskussion

WILLI WOLF

Beurteilungsmodelle bei Klassenarbeiten. Bestandsauf-
nahme schulischer Praxis in der Sekundarstufe I 115

V. Besprechungen

ANDREAS KRAPP	ARBEITSGRUPPE SCHULFORSCHUNG: Leistung und Versagen 131
ANDREAS KRAPP	H. GERHARD BEISENHERZ et al.: Schule in der Kritik der Betroffenen 131
PETER GSTETTNER	WOLFGANG KLAFKI et al.: Schulnahe Curriculumentwicklung und Handlungsforschung 135
JÜRGEN DIEDERICH	PETER FAUSER/KLAUS J. FINTELMANN/ANDREAS FLITNER (Hrsg.): Lernen mit Kopf und Hand 141

VI. Dokumentation

Pädagogische Neuerscheinungen 145

Hinweise zur äußeren Form der Manuskripte für die „Zeitschrift für Pädagogik“ 147

Vorschau auf Heft 2/84

Themenschwerpunkt „Jugendprobleme“, u. a. mit folgenden Beiträgen:

- Helga Bilden/Angelika Diezinger: Individualisierte Jugendbiographie?
- Walter Hornstein: Neue soziale Bewegungen und Pädagogik
- Marianne Kieper: Selbstkontrolle, Selbstbehauptung, Sexualität

Zu den Beiträgen in diesem Heft

EGON SCHÜTZ: *Einige Überlegungen zur Fragwürdigkeit systematischer Pädagogik*

Der vorgestellte Gedankengang geht aus von der Exposition des Problems einer „allgemeinen“ Pädagogik unter Bedingungen jener zunehmenden theoretischen und praktischen Differenzierung des Erziehungsphänomens, die offenbar Momente grundlegender Allgemeinheit im Bereich der Pädagogik weitgehend in Frage stellt. Im gedanklichen Gegenzug gegen die spezialistischen Auflösungstendenzen wird der Versuch unternommen, durch Aufweis einer existentialen Strukturierung von Praxis als „Entwurf“ anthropologische Elementaria freizulegen, die sich gleichsam unterhalb der eingespielten Theorie-Praxis-Differenz durchhalten. Genannt und umrissen werden: Zeitlichkeit, Räumlichkeit, Leiblichkeit, Sprachlichkeit und Vernunft. In der Folge dieses Ansatzes wäre eine Allgemeine Pädagogik allerdings primär nicht mehr durch ein System der Wissenschaften oder eine Systematik professioneller Tätigkeitsfelder zu gewinnen, sondern in einer erziehungsphilosophischen Archäologie der Existenz und deren fragendem Rückbezug auf positives Wissen und spezialisierte Praxis.

ROLF HUSCHKE-RHEIN: *Über die Zukunft der Allgemeinen Pädagogik. Systematische und systemökologische Überlegungen*

Beide Typen Allgemeiner Pädagogik, die sich im modernen Hochschulbereich entwickelt haben, die kulturistische „Allgemeine Pädagogik“ und die ahistorische „Systematische Pädagogik“, können den Aufgaben zukünftiger Allgemeiner Pädagogik nicht mehr genügen, da in beiden die Modalitäten der „Zukunft“ und der „Möglichkeit“ fehlen (Abschnitt 1). Am Thema „Allgemeinbildung und Berufsbildung“ wird der Widerstreit zwischen den Kategorien der Möglichkeit und der Wirklichkeit bzw. zwischen denen der Zukunft (des Menschen) und der Faktizität (der beruflichen Realität) auf prägnante Weise verdeutlicht (Abschnitt 2). Erst eine allgemeine Theorie der Natur, wie sie sich derzeit in ökologischer Systemtheorie, Kosmologie und theoretischer Physik abzuzeichnen beginnt, wird aus den aufgezeigten Problemen herausführen können; sie ermöglicht eine Neubestimmung des Verhältnisses von Mensch, Gesellschaft und Natur und damit auch eine veränderte Praxis (Abschnitt 3).

HEINZ-ELMAR TENORTH: *Berufsethik, Kategorialanalyse, Methodenreflexion. Zum historischen Wandel des „Allgemeinen“ in der wissenschaftlichen Pädagogik*

Die Abhandlung untersucht Muster des „Allgemeinen“ in zwei Etappen der Pädagogik: (1) „Berufsethik“ in der Schulmännerpädagogik des 19. Jahrhunderts, die ihre theoretischen und praktischen Gewißheiten über fraglos geltende Normen und Institutionen bezog und Wissen noch eindeutig in Handlungskontexte integrieren und auf sie hin auslegen konnte; (2) „Kategorialanalyse und Methodenreflexion“ in der Erziehungswissenschaft, die ihre Gewißheiten selbst erzeugen und mit der Differenz theoretischer und

praktischer Geltung leben muß. Für die geisteswissenschaftliche Pädagogik wird zudem gezeigt, daß in dieser „Theorie von und für Praxis“ Verbindlichkeitserwartungen des Handelns zum Schaden der Erkenntnis die Theoriepraxis bestimmen. Für die Systematik allgemeiner Erziehungswissenschaft werden schließlich Wissensformen und Relationen von Analyse und Forschung unterschieden, die bisher z. T. konfundiert oder „philosophisch“ übersprungen werden.

HANS BRÜGELMANN: Lesen- und Schreibenlernen als Denkentwicklung. Voraussetzungen eines erfolgreichen Schrifterwerbs

In diesem Bericht aus dem Bremer Projekt „Kinder auf dem Weg zur Schrift“ werden die Voraussetzungen eines erfolgreichen Schrifterwerbs unter Rückgriff auf empirische Untersuchungen aus verschiedenen Ländern diskutiert. Schwierigkeiten beim Schrifterwerb lassen sich danach nicht auf einen bestimmten Bereich eingrenzen. Weder sinnesspezifische Funktionsstörungen noch allgemeine intellektuelle Operationen oder die Entwicklung der gesprochenen Sprache können die Streuung der Lese- und Rechtschreibleistungen nach dem 1. bzw. 2. Schuljahr befriedigend über die Gesamtgruppe hinweg erklären. Es gibt alternative Ursachen für jeweils unterschiedlich ausgeprägte Lese-/Rechtschreibschwächen. Eine besondere Bedeutung kommt dabei den konkreten Erfahrungen mit Schrift zu, die Kinder vor und neben der Schule sammeln. Über einzelne Kenntnisse und Fertigkeiten wie Buchstaben- oder Wortkenntnis hinaus sind die gedanklichen Vorstellungen der Kinder über die soziale Funktion und über den technischen Aufbau der Schrift wichtig. Diese kindlichen Konzepte sind zwar naiv, aber nicht willkürlich, sondern in sich geordnet und helfen, auf dem jeweiligen Entwicklungsstand subjektiv sinnvoll mit Schrift umzugehen. Einheitsfibeln, durch die alle Schulanfänger einer Klasse sozusagen im Gleichschritt marschieren müssen, werden dieser Situation nicht gerecht. Mehr noch: Lesen- und Schreibenlernen läßt sich nicht als passives Abspeichern von stückweise verabreichten Wissensbausteinen und als Addition isolierter Teilleistungen erklären; das Kind ist auf persönlich bedeutsame Erfahrungen aus dem aktiven Umgang mit Schrift angewiesen, um seine individuellen Vorstellungen fortentwickeln zu können.

MECHTHILD DEHN: Lernschwierigkeiten beim Schriftspracherwerb. Kriterien zur Analyse des Leselernprozesses und zur Differenzierung von Lernschwierigkeiten

Die Forschung hat sich dem Prozeß des Schriftspracherwerbs bisher kaum zugewandt. Das „Wissen“ über den Schriftspracherwerb orientiert sich an der Analyse der Struktur der geschriebenen Sprache, an Untersuchungen über das Lesen und Schreiben des Könners oder an den manifest gewordenen Lernschwierigkeiten älterer Schüler. (Diese Beschränkung ist vermutlich nicht zuletzt methodisch begründet.) Die Autorin versucht die forschungsmethodischen Probleme des Schriftspracherwerbs zu lösen – sowohl bei der Konstruktion der Beobachtungssituation und der lehrgangsbezogenen Aufgabenstellung als auch bei der Auswertung der Befunde. Die Kriterien, die sie für eine Interpretation schulischer Lernprozesse zur Diskussion stellt, sind denkpsychologisch und kognitions-theoretisch begründet. Sie können dazu dienen, Lernschwierigkeiten frühzeitig zu erkennen. Schließlich lassen die Ergebnisse gängige Vorstellungen über die Beziehung von Lehren und Lernen durchaus als fragwürdig erscheinen.

WILLI WOLF: *Beurteilungsmodelle bei Klassenarbeiten. Bestandsaufnahme schulischer Praxis in der Sekundarstufe I*

Der Beitrag informiert über Beurteilungsmodelle, die Lehrer bei der Bewertung schulischer Klassenarbeiten verwenden – eine bisher weitgehend vernachlässigte Frage. Zunächst wird über die Ergebnisse einer explorativen empirischen Studie berichtet, in der 51 Lehrer aus unterschiedlichen Schulformen und mit unterschiedlichen Fächern zu ihrer Klassenarbeitspraxis interviewt wurden, und es wird versucht, die Existenz eher impliziter Beurteilungsmodelle nachzuweisen. Die Ergebnisse werden im Hinblick auf Validitäts- und Reliabilitätsforderungen diskutiert. Gesondert wird auf die Möglichkeiten eingegangen, auch die Beurteilung von Deutschaufsätzen für Schüler transparent zu gestalten.

Contents and Abstracts

Essay:

KONRAD WÜNSCHE: *Ludwig's First Months in School. A Conversation Between Educators* 1

Topic: Educational Theory

EGON SCHÜTZ: *Some Reflexions on the Questionable Nature of Systematic Pedagogics* 17

The train of thought presented in this study is rooted in the exposition of the problematic nature of a "general" educational theory under the conditions of the increasing theoretical and practical differentiation of the educational system, conditions which seem to make it impossible to talk about fundamental factors of the attribute "general" in the field of education. In a theoretical countermove against the trend towards specialization and disintegration an attempt is made to point out fundamental anthropological factors which can be called "general" despite the common distinction between theory and practice; these factors are to be revealed by means of an existentialist construction of praxis as a design ("Entwurf"). The following factors will be discussed in outline: temporality, spatiality, corporality, human speech, and reason. As a consequence of this approach, however, a general educational theory could no longer be primarily derived from either a system of the sciences nor from a system of the professional fields of activity, but rather from an educational-philosophical archeology of existence and from the inquiring reflexion on positive knowledge and specialized praxis held therein.

ROLF HUSCHKE-RHEIN: *On the Future of the General Theory of Education. Systematic Reflexions on an Ecological Theory of Systems* 31

Neither of the two forms of general educational theory which have been developed in the modern academic realm, "culturistic" "General Pedagogics" and ahistorical "Systematic Pedagogics", can meet the demands of the future general educational theory, since both lack the two modalities of "future" and of "potentiality" (Part 1). It is in the topic of general versus vocational education that the antagonism between these two categories of potentiality and reality, and, respectively, between those of future (of mankind) and factuality (of the vocational reality) becomes clearly apparent (Part 2). Only a general theory of nature such as is, at present, taking shape in the ecological theory of systems, in cosmology, and in theoretical physics, will solve the problems outlined here. Such a theory will allow a redefinition of the interrelation of man, society, and nature, and, therefore, also a changed conduct of life (Part 3).

HEINZ-ELMAR TENORTH: *Professional Ethics, Categorical Analysis, Methodological Reflexions. Reflexions on Historical Change in the Concept of the Attribute "General" in Educational Science* 49

In this essay patterns of the attribute "general" in two phases of the history of pedagogics are examined: (1) "professional ethics" in the 19th century pedagogics of educationalists who derived their theoretical and practical certainties from unquestioned norms and institutions, and who had no problem in integrating their knowledge into everyday practice; (2) "categorical analysis and methodological reflexion" in educational science which has to develop its own certainties and that has to live with the difference between theoretical and practical validity. For the "humanistic" pedagogics it is shown that in this "theory of and for educational practice," the expectations of binding regulations for the practitioner determine theorizing to the detriment of the acquisition of knowledge. Finally, it is differentiated between different knowledge forms and relations of analysis and research in the system of general educational theory which have, until now, been either confused or overlooked.

Topic: The Acquisition of Reading and Writing Skills

HANS BRÜGELMANN: *Learning How to Read and Write as a Process of Intellectual Development* 69

In this account of the research project "Children on their way to written language," carried out in Bremen, the prerequisites for a successful acquisition of literary skills are discussed, drawing on empirical research from various countries. According to these studies, difficulties in acquiring literary skills cannot be limited to a single area: neither sensory malfunction, nor general intellectual operations, nor the development of oral language can explain satisfactorily the variation in the reading and writing achievements of the total population after the first or second year in school. There are alternative explanations for distinct reading or writing problems. Of special significance are the concrete experiences with written language that the children have had before they start school and outside of the classroom. Not only isolated accomplishments and skills, such as knowing single letters or words, are important, but also the intellectual conceptions that the children have of the social function and of the technical structure of written language. These concepts formed by the child may be naive, they are, however, not arbitrary but rather well ordered, and they help the child to deal with written language in a way that is subjectively meaningful at a given developmental stage. Therefore, primers which force a uniform progression upon all children of one classroom, do not meet the situation. Moreover, the acquisition of reading and writing skills cannot be interpreted either as a passive storing of bits of information or as the adding up of partial skills. In order to further the development of individual concepts the child is dependent on personally significant experiences in the active approach to written language.

MECHTHILD DEHN: *Difficulties in Primary Reading. Criteria for Analysing the Process of Learning to Read and for Differentiating the Difficulties of Learning* 93

Research into primary reading has up till now hardly faced this actual process: our knowledge is orientated on the one hand towards the analysis of the structure of written language, on the other hand towards examinations of the reading and writing of the advanced learner or towards the learning difficulties of older pupils. Methodological problems are probably the cause for this deficiency. The study is concerned with the methodological problems of research – in constructing the setting of observation and the curricular task as well as in the analysis of the findings. The criteria which are set forth here in order to be discussed for an interpretation of learning processes in school, are well-founded in the psychology of thinking and in the theory of cognition. They can help to discover difficulties in learning at an early stage. The results do cast some doubt on current ideas about the relationship between teaching and learning.

Discussion

WILLI WOLF: *Conceptual Models of Grading Nonstandardized Tests. An Evaluation of Grading Practice in Junior High Schools* 115

This paper gives information about evaluation models used by teachers in evaluating examination papers written at school. This problem has been almost wholly neglected so far. The interpretation is based on the results of an exploratory empirical study in which 51 teachers from various school types and for various subjects were interviewed on their evaluation practice. An attempt is made to prove the existence of more implicit evaluation models. The results are discussed with regard to the demands of validity and reliability. In addition, there is also a discussion of the possibility of making the evaluation of essays in the mother tongue (German) transparent to the students.

Book Reviews 131

New Books 145

Beurteilungsmodelle bei Klassenarbeiten

Bestandsaufnahme schulischer Praxis in der Sekundarstufe I

1. Einleitung

1.1. Vorbemerkungen

Bei dem nachfolgenden Beitrag handelt es sich um einen Teil eines Berichtes aus dem Projekt „Klassenarbeiten“¹, mit dem der Autor eine Bestandsaufnahme schulischer Klassenarbeitspraxis in der Sekundarstufe I versuchte. Der Anlaß für dieses Projekt lag vor allem in den folgenden Sachverhalten:

- Eine Durchsicht gängiger, neuerer und älterer Monographien und Sammelbände aus dem Bereich schulischer Leistungsbeurteilung² läßt nicht nur im Hinblick auf die durch Erlasse und Verordnungen³ geregelten Anforderungen an schulische Leistungsbeurteilungen, sondern auch im Hinblick auf die konkret in Schulen ablaufenden Beurteilungsprozesse u. a. folgendes Defizit deutlich werden: es fehlen wissenschaftlich und praktisch begründete Handlungsmuster für die Beurteilung schriftlicher Schülerleistungen in Form von Klassenarbeiten.
- Dieses Defizit korrespondiert in der Literatur häufig mit der Ablehnung konventioneller Klassenarbeiten, da diese den Anforderungen nach Reliabilität und Validität nicht genügten, und mit der Forderung, konventionelle Klassenarbeiten durch Tests, vor allem durch informelle Tests, zu ersetzen. Vorschläge zur Verbesserung der Klassenarbeitspraxis lassen sich hingegen nur in seltenen Fällen finden.
- Im erziehungswissenschaftlichen Studium für Lehrer, so lassen sich die Erfahrungen des Autors aus einschlägigen Lehrveranstaltungen und zahlreichen Staatsexamensprüfungen zusammenfassen, liegt bei der Mehrzahl der Studenten eine Einstellung zur schulischen Leistungsbeurteilung vor, die sich überwiegend durch eine völlige Ablehnung der Notengebung in der Schule und/oder durch eine Hilflosigkeit gegenüber diesbezüglichen Anforderungen des späteren Schulalltags charakterisieren läßt.
- Gespräche mit Lehrern über das Problem der Bewertung schriftlicher Schülerarbeiten, die in Hessen auf örtlicher und regionaler Ebene geführt wurden, ergaben folgende wichtige Ergebnisse:
 1. Nach wie vor ist die konventionelle Klassenarbeit die Hauptform der Überprüfung schriftlicher Schülerleistungen. Informelle schulklassen- oder lerngruppenbezogene Tests werden selten und überregionale normorientierte Tests fast gar nicht verwendet.

-
- 1 Das Projekt „Klassenarbeiten“ wurde dankenswerterweise durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft ermöglicht.
 - 2 FRIEDRICH, L. / KOHLER, K. (Hrsg.): Zeugnisnoten und Numerus Clausus. Kronberg 1975.
HELLER, K. (Hrsg.): Leistungsbeurteilung in der Schule. Heidelberg 1978.
INGENKAMP, K. (Hrsg.): Die Fragwürdigkeit der Zensurengebung. Weinheim 1972.
INGENKAMP, K. (Hrsg.): Schüler- und Lehrerbeurteilung. Weinheim und Basel 1977.
KLAUER, K. J., u. a.: Lehrzielorientierte Tests. Düsseldorf 1972.
KLEBER, E. W., u. a.: Beurteilung und Beurteilungsprobleme. Weinheim und Basel 1976.
KUTSCHER, J. (Hrsg.): Beurteilen oder verurteilen. München und Wien 1977.
STEPHAN, E. / SCHMIDT, W. (Hrsg.): Messen und Beurteilen von Schülerleistungen. München und Wien 1978.
 - 3 S. z. B. die in Hessen gültige „Verordnung über schriftliche Arbeiten“ vom 3. 7. 1978 (Amtsblatt des Hessischen Kultusministers, 1978, S. 328 ff.).

2. Für die Konzipierung und Bewertung von Klassenarbeiten sind den Lehrern aus dem Studium höchst selten Modelle von Handlungsanleitungen (Handlungshilfen) bekannt. Vereinfachend läßt sich sagen, daß Lehrer im Verlaufe ihres Berufslebens *individuell* Handlungskonzepte für ihre *eigene* Klassenarbeitspraxis entwickeln. Diese individuellen Konzepte werden im Regelfall nicht mit den Kollegen diskutiert, geschweige denn, daß sie in der Schule oder schulübergreifend abgestimmt würden. Momente einer Verbreitung solcher Konzepte zeigen sich allenfalls und ansatzweise in der Referendarausbildung, wenn der „erfahrene“ Lehrer dem „Berufsanfänger“ seine eigenen Handlungskonzepte zu vermitteln versucht.

1.2. Generelle Zielsetzung des Projektes

Geht man von der plausiblen und nicht diskriminierend gemeinten Voraussetzung aus, daß es in der Fähigkeit, Schülerleistungen zu beurteilen, Unterschiede zwischen Lehrern gibt – d. h., daß es Lehrer gibt, deren Beurteilungspraxis allseitig (bei Schülern, Kollegen und Eltern) weitgehend akzeptiert wird, und solche, die hier Schwierigkeiten haben –, dann kann es eine lohnende Aufgabe sein, die bewährte Beurteilungspraxis der ersteren zu erfassen, zu systematisieren und zu Handlungsanleitungen/-hilfen umzuformen.

In diesem Versuch wird ein Ansatz gesehen, eine wichtige Berufsfunktion des Lehrers zu beschreiben und die Ergebnisse einerseits für Aus- und Weiterbildungsprozesse nutzbar zu machen und andererseits der weiteren wissenschaftlichen Diskussion zuzuführen.

Dieses generelle Ziel der Bestandsaufnahme oder Inventarisierung schulischer Klassenarbeitspraxis aus Lehrersicht läßt sich unter der Voraussetzung, daß die jeweilige Schulform und das Schulfach eine Rolle spielen können, durch folgende Forderungen kennzeichnen:

- a) Bei der Inventarisierung muß in Abhängigkeit von Schulformen und -fach versucht werden, Homogenität und Heterogenität der Klassenarbeitspraxis zu erfassen. Leitfragen sind somit: Welche Verhaltensweisen treten bei *Gruppen* von Lehrern auf, werden von *Gruppen* von Lehrern als wichtig angesehen? Welche Verhaltensweisen treten nur bei *einzelnen* Lehrern auf, werden nur von *einzelnen* Lehrern als wichtig angesehen?
- b) Klassenarbeiten lassen sich – bezogen auf die erforderlichen Lehreraktivitäten – als Prozeß darstellen, der in vier Phasen abläuft:
 1. Planen/Entwurf der Klassenarbeit durch den Lehrer,
 2. Schreiben der Klassenarbeit in der Klasse,
 3. Beurteilung der Klassenarbeit durch den Lehrer und
 4. Rückgabe der Klassenarbeit an die Klasse.
- c) Im Hinblick auf diese Phasen gilt es vor allem, konsistente Praxiserfahrungen des Lehrers zu erfassen.

1.3. Planung und Durchführung des Projektes

Entsprechend der offenen Zielsetzung des Projektes wurde es als zweckmäßig angesehen, für die Bestandsaufnahme einen methodisch weitgehend offenen Zugang zur Datengewinnung zu wählen. Zwar wäre für einen oder wenige Fälle das Verfahren der teilnehmenden Beobachtung optimal gewesen, bei einer größeren Lehrergruppe ist diese Methode jedoch schon aus Zeitgründen nicht realisierbar. Als brauchbare Alternative boten sich offene

Interviews auf der Basis eines vorstrukturierten Interviewleitfadens an. Die Grobstruktur dieses Interviewleitfadens entspricht den in Abschnitt 1.2. genannten vier Phasen, die um den Komplex „Sonstiges“ ergänzt wurden. Die einzelnen Interviewvorgaben (Anreize) innerhalb jedes der fünf Komplexe oder Leitthemen waren überwiegend an möglichen Lehrertätigkeiten, Lehrereinstellungen und Lehrerzielsetzungen orientiert. Die Interviews sollten mit Tonband aufgezeichnet werden; ihre Dauer wurde auf rund eine Zeitstunde veranschlagt.

Der Untersuchungsplan (Interviewplan) mußte so angelegt werden, daß schulform- und schulfachspezifische Auswirkungen sichtbar werden konnten. 64 Lehrer sollten in das Projekt einbezogen werden: jeweils 16 in jedem Fach und in jeder Schulform und jeweils vier für jede Kombination von Fach und Schulform (vgl. Tabelle 1).

Tabelle 1: Interviewplan nach Schulform und -fach

	Förder- stufe	Haupt- schule	Real- schule	Gymna- sium	Ins- gesamt
Deutsch	4 (4)	4 (3)	4 (3)	4 (2)	16 (12)
Englisch	4 (3)	4 (3)	4 (3)	4 (3)	16 (12)
Mathematik	4 (3)	4 (3)	4 (4)	4 (4)	16 (14)
Naturwissenschaft	4 (3)	4 (3)	4 (3)	4 (4)	16 (13)
Insgesamt	16 (13)	16 (12)	16 (13)	16 (13)	64 (51)

Die Auswahl der Schulen, die aus Zeit- und Finanzgründen auf den Landkreis Marburg-Biedenkopf beschränkt war, sollte mit Hilfe des zuständigen staatlichen Schulamtes erfolgen; die Auswahl der Lehrer hingegen über die Schulleiter, die entweder selbst oder zusammen mit der Lehrerkonferenz die Namen von jeweils vier erfahrenen Kolleginnen/Kollegen mitteilen sollten. Dieses Auswahlverfahren ließ sich ohne größere Schwierigkeiten realisieren⁴.

Die von den Schulen benannten Lehrer wurden jeweils individuell angeschrieben und um ihre Mitarbeit gebeten, wobei ihnen u. a. der Projektentwurf und ein Exemplar des zwischenzeitlich getesteten Interviewleitfadens mitgeschickt wurden. Die faktische Interviewstichprobe (s. die in Tabelle 1 eingeklammerten Zahlen) wich aus unterschiedlichen Gründen⁵ von der geplanten Stichprobe ab. Es konnten schließlich nur 51 Lehrer interviewt werden⁶.

4 Es waren insgesamt zwölf Schulen an dem Projekt beteiligt.

5 Die wichtigsten Gründe waren:

- Der Schulleiter konnte keine Namen vorschlagen, da sich schon vorab nicht hinreichend viele Lehrer zur Teilnahme bereit erklärten.
- Nach Kenntnis des Projektentwurfes und des Interviewleitfadens lehnten einige der angeschriebenen Lehrer ihre Teilnahme – vorwiegend aus zeitlichen Gründen – ab.

6 In der überwiegenden Zahl der Fälle wurden ausschließlich konventionelle Klassenarbeiten geschrieben; informelle Tests i. e. S. tauchten vor allem in den naturwissenschaftlichen Fächern und in den unteren Klassen auf.

Die Interviews wurden im November und Dezember 1981 von einem Interviewer durchgeführt und fanden etwa zu zwei Dritteln in Schulen (vor- und nachmittags) und zu einem Drittel in Privatwohnungen statt. Sie dauerten von knapp einer Stunde bis zu maximal drei Stunden, wobei die in Schulen am Vormittag durchgeführten Interviews im Schnitt kürzer ausfielen als die übrigen Interviews.

1.4. Auswertung der Interviews

34 der insgesamt 51 Interviews wurden transkribiert und führten zu Texten mit einem Umfang von 18 bis 45 Schreibmaschinenseiten⁷. Die Auswertung vollzog sich dann in folgenden Schritten:

- 16 der transkribierten Interviews, je eines für jede Kombination von Fach und Schulform, wurden in Form von Inhaltsanalysen ausgewertet. Diese Auswertung orientierte sich sehr stark an den jeweiligen Lehreraussagen. Der erste Arbeitsschritt führte bei jedem Interviewanreiz zu einem Bündel teilweise hierarchisch aufeinander aufbauender Auswertungskategorien. Dieser Prozeß wurde gemeinsam von zwei Codierern durchgeführt.
- Nach dieser Vorarbeit erfolgte die Auswertung der übrigen transkribierten und der restlichen 17 nur auf Tonband vorliegenden Interviews, wobei noch knapp die Hälfte aller dieser Interviews von den Codierern gemeinsam ausgewertet wurde. Fast jedes Interview führte bei vielen Interviewanreizen zur Ergänzung der jeweiligen Kategoriensysteme.
- Die Ergebnisse der Auswertung wurden für jede Vorgabe jedes Interviews, getrennt nach Schulformen und -fächern, auf Auswertungsformulare notiert, die auch eine Auflistung der jeweiligen Auswertungskategorien enthielten⁸.
- Im Hinblick auf die Zielsetzung der Untersuchung wurden besondere Reliabilitätskontrollen als nicht notwendig angesehen⁹.
- Die Interpretation der Ergebnisse erfolgte anhand der Auswertungsformulare und unter Hinzuziehung der 34 transkribierten Interviews. Die erforderlichen Häufigkeitsauszählungen erfolgten per Hand.

1.5. Drei Hinweise zum Stellenwert der Ergebnisse

Die im Rahmen des oben geschilderten methodischen Vorgehens gewonnenen Ergebnisse unterliegen drei Einschränkungen, die jeweils bei der Interpretation und Verwendung der Ergebnisse zu bedenken sind.

- a) Entsprechend der Zielsetzung des Projektes war keine repräsentative Untersuchung, z.B. repräsentativ für alle Lehrer des ausgewählten Landkreises, beabsichtigt. Die Ergebnisse dürfen somit auch nicht in dieser Hinsicht verallgemeinert werden.

⁷ Aussagendichte und Ergiebigkeit der Interviews variierten deutlich. Ursachen dieser Variation werden vor allem in den unterschiedlichen Durchführungsbedingungen (Interview in 45 Minuten plus Pause am Schulvormittag im Lehrerzimmer versus Interview am Nachmittag im privaten Lehrerarbeitszimmer und ohne unmittelbar spürbaren Zeitdruck) und in dem unterschiedlichen Ausmaß der Vorbereitung auf das Interview gesehen (einige Lehrer hatten den Interviewleitfaden vorher gründlich durchgearbeitet, andere sich hingegen nur einen groben Überblick verschafft).

⁸ Der Codierprozeß wurde in zwei Phasen im Sommer 1982 durchgeführt.

⁹ Bei der Auswertung und der Interpretation wurden – anders als beim linear verlaufenden Codierprozeß – zu einem Interviewanreiz jeweils alle Ergebnisse aus allen transkribierten Interviews parallel gelesen und mit den Codiereinträgen verglichen. Abweichungen tauchten nur in seltenen Fällen auf, relativ häufig jedoch ergaben sich noch Zusätze zu den ursprünglichen Codierungen.

- b) Die Klassenarbeitspraxis in Schulen ist nicht allein durch die Erfassung der Lehrererfahrungen in Form von Interviews beschreibbar. Es fehlen zumindest die Schülererfahrungen als ergänzende Komponente. So kann sich die Wahrnehmung der Lehrerpraxis durch die Schüler von der Selbstwahrnehmung der Lehrer deutlich unterscheiden. Auch hier werden Einschränkungen im Hinblick auf die Aussagekraft der Ergebnisse zu berücksichtigen sein.
- c) Aus der oft feststellbaren Tatsache, daß ein Lehrer in seinem Interview einen bestimmten Sachverhalt nicht anspricht, kann nicht geschlossen werden, daß dieser Sachverhalt in seiner Praxis unwichtig sei oder überhaupt nicht auftrete. Wenn z. B. acht von zwölf interviewten Deutschlehrern angeben, daß sie bei Deutschaufsätzen die Rechtschreibung korrigieren, so heißt das nicht, daß bei den restlichen vier Deutschlehrern keine Rechtschreibkorrektur erfolgt; sie fanden es vielleicht, da trivial, nicht erwähnenswert.

2. Welche Beurteilungsmodelle lassen sich in der Schulpraxis vorfinden?

2.1. Fazit der Bestandsaufnahme

Vor der Wiedergabe und Diskussion einzelner Ergebnisse (s. Abschnitt 2.2.) soll zunächst der Versuch einer Systematisierung der diesbezüglichen Interviewäußerungen erfolgen. In diese Systematisierung sind alle die Sachverhalte einbezogen, die mindestens einmal in einem Interview explizit angesprochen wurden. Bei dieser Bestandsaufnahme wird versucht, die einzelnen Sachverhalte durch Zitate aus den Interviews zu konkretisieren und zu veranschaulichen.

Nach Meinung der interviewten Lehrer lassen sich bei den Beurteilungsmodellen zunächst zwei grundlegende Typen unterscheiden, die man als „Richtlinien-Modell“ und als „Maßstab-Modell“ bezeichnen kann.

2.1.1. Richtlinien-Modelle

Richtlinien-Modelle werden im Bereich solcher Schülerleistungen verwendet, die von den betroffenen Lehrern als schwer quantifizierbar eingeschätzt werden. Sie stellen in der Regel eine nur grobe Festlegung der vom Lehrer erwarteten Schülerleistung dar:

- (1) „Ich darf das an einem Beispiel erläutern. Wir haben in der letzten Woche in der R 9 zur ‚Judenbuche‘ einen Aufsatz mit dem Thema geschrieben: ‚Charakterisiere die Figur des Friedrich Merkel, und begründe, warum dieser zum Mörder wurde.‘ Da habe ich schon konkrete Vorstellungen, was ich hören will. Und ich glaube, daß man das bei einem so formulierten Thema auch festlegen kann: das möchte ich vom Schüler hören, er soll zeigen, daß er die Figur charakterisieren kann; das ist ja auch bekannt, ist im Grunde teilweise reproduzierend. Aber dann kommt da noch der zweite Teil des Themas: ‚Warum wurde er so, wie er geworden ist?‘ Das ist dann das, was der Schüler selbst entwickeln muß. Bei diesem Punkt habe ich dann aber auch schon meine Vorstellungen. Die Unterschiede – wenn man an die Bewertung denkt – ergeben sich daraus, wie es dargestellt wird, wie es vermittelt wird, wie es strukturiert ist und natürlich auch, ob der Schüler überhaupt zu dieser Frage gekommen ist oder ob er das Thema verfehlt hat... Mit der Formulierung des Themas wird an sich vorher festgelegt, was intendiert ist.“ (RS, D)¹⁰

10 Zur Kennzeichnung von Schulform und Fach werden folgende Abkürzungen verwendet:

FÖ: Förderstufe
HS: Hauptschule

RS: Realschule

GY: Sekundarstufe I des Gymnasiums

D: Deutsch

E: Englisch

M: Mathematik

N: naturwissenschaftliches Fach

Im Detail erfolgt die Ermittlung der Abweichung von den Erwartungen in Formen, die man als eher „subtraktiv“ oder als eher „additiv“ bezeichnen kann. Damit ist folgendes gemeint. Die subtraktive Form ist an fehlerhaften Sachverhalten oder an Mängeln orientiert. Bei ihr liest der Lehrer die Klassenarbeit eines Schülers und versucht, Verstöße gegenüber einer Erwartung oder Norm zu registrieren. Lassen sich keine solchen Verstöße feststellen, so stellt die Arbeit eine sehr gute Leistung dar; mit wachsender Anzahl sinkt die Bewertung der Leistung.

- (2) „Ich arbeite eigentlich eher nach dem Abstrichverfahren. Ich sehe also, was ein Schüler alles hätte machen können, und dann streiche ich ab, wieviel von seiner Arbeit dem eigentlich noch entspricht. Dann bleibt eben mehr oder minder viel übrig. Es ist ja ganz klar, wenn ein Schüler überwiegend Handlungen nacherzählt, daß er dann nicht die Person vollständig beschreiben kann... Ja, das könnte man so beschreiben, daß man von einem Idealbild ausgeht.“ (GY, D)

Die additive Form ist an richtigen Sachverhalten orientiert: Welche Sachverhalte hat der Schüler im Hinblick auf eine Erwartung oder Norm richtig dargestellt? Durch additive Zusammenfassung dieser richtigen Sachverhalte gelange man dann bezüglich der Erwartung zu einer Beurteilung.

- (3) „Da konnte man ein Raster anlegen und sagen, nach den Aussagen, die in dem Kindergespräch enthalten waren, hätten die Schüler die folgenden Sachverhalte bei der Beschreibung liefern müssen.“ (FÖ, D)

Richtlinien-Modelle führen meist unmittelbar zu Noten, die als Versuch der Quantifizierung des Bewertungseindrucks zu verstehen sind.

Da nur in seltenen Fällen für eine Klassenarbeit nur eine Erwartung existiert, im Regelfall liegt immer ein Erwartungsbündel (etwa inhaltliche und sprachliche Gestaltung im Deutschaufsatz) vor, stellt sich hier zusätzlich ein Gewichtungproblem – vgl. Zitate 4 und 6 –, bei dem auch die Festlegung der Note „ausreichend“ eine Rolle spielen kann (5)¹¹.

- (4) „Da würde ich immer sagen 60 zu 40, denn das Inhaltliche ist doch wichtiger als das Formale. Wenn da mal einer die Zeit wechselt oder wirklich einmal eine wörtliche Rede da reinbringt, ist das nicht so gravierend, als wenn er jetzt den Inhalt nicht verstanden hat.“ (HS, D)
- (5) „Was ist der Kern der Geschichte, die da herausgearbeitet werden soll? Wenn einer den Kern der Geschichte erfaßt hat, dann mache ich das so, daß ich die Arbeit mit ‚ausreichend‘ bewerte.“ (HS, D)
- (6) „Es gibt Englisch-Lehrer, die sagen, ich lehre hier Englisch, mich interessiert nur das Sprachliche, das Inhaltliche sollen sie sich im Deutschunterricht oder – vielleicht war es ein naturwissenschaftliches Thema – im Naturwissenschaftsunterricht benoten lassen. Ich bin dagegen, denn ich meine, auch hier kann ich, wenn ich das Ganze rein inhaltlich betrachte, erkennen, ob ein Schüler die Sprache so beherrscht, daß er die Inhalte aus einer bestimmten Textvorgabe heraus überhaupt erkennen kann, komplex oder weniger komplex. Das sagt ja auch etwas zur Sprachbeherrschung aus. Ich meine also, daß es durchaus berechtigt ist, die inhaltlichen Gesichtspunkte gleichberechtigt neben die sprachlichen zu setzen.“ (RS, E)

2.1.2. Maßstab-Modelle

Den „*Maßstab-Modellen*“, d. h. dem anderen Typus der Beurteilungsmodelle, liegt im Regelfall eine Bewertung nach Punkten, die ein Schüler in einer Klassenarbeit erreicht

¹¹ Werden zu einem Sachverhalt mehrere Passagen zitiert, so stammen diese jeweils aus unterschiedlichen Interviews.

hat, oder nach den Fehlern, die er gemacht hat, zugrunde. So ergeben sich zwei Grundformen: Punkte-Modelle und Fehler-Modelle. Sie werden im Bereich solcher Schülerleistungen verwendet, die als leichter zu quantifizieren gelten, und repräsentieren in dieser Hinsicht detailliertere Formen der erwarteten Schülerleistung.

2.1.1.1. Punkte-Modelle

Bei Punkte-Modellen gibt es zwei Ansätze, die sich als additives (A) und als subtraktives Punkte-Modell (B) beschreiben lassen.

A) „*Additives Punkte-Modell*“ besagt, daß der Schüler bei einer bestimmten Aufgabe eine – im Regelfall – vorher festgelegte Punktzahl erreichen kann. Die Summe dieser Aufgabenpunktzahlen ergibt die maximal mögliche Gesamtpunktzahl. Eine individuelle Schülerleistung läßt sich dann im Hinblick auf diese Gesamtpunktzahl beurteilen. Additive Punkte-Modelle gehen im Schwerpunkt von richtigen Sachverhalten aus; sie korrespondieren in dieser Hinsicht der additiven Form bei den Richtlinien-Modellen.

- (7) „Das wird vorher festgelegt, und zwar wird pro Aufgabe eine bestimmte Anzahl von Punkten oder Bewertungseinheiten vorgegeben.“ (GY, N)

Bei der Verwendung additiver Punkte-Modelle können noch folgende Entscheidungen bedeutsam sein:

- a) Wird eine Aufgabe insgesamt betrachtet, oder wird sie noch in Teilleistungen zerlegt? ((8) und (9))
 - b) Werden nur richtige Sachverhalte bewertet, oder führen falsche Sachverhalte auch zu Punktabzügen? (9)
 - c) Gibt es für jede Aufgabe die gleiche Punktzahl, oder erhalten z.B. schwerere Aufgaben eine höhere Punktzahl? ((8) und (10) bis (12))
- (8) „In den letzten Schuljahren habe ich in der Regel die Mathematikarbeiten der Mittelstufe mit Bewertungseinheiten, mit Punkten, belegt; also eine Punktwertung durchgezogen. Diese Festlegung der Punkte geschieht in der Regel schon bei der Planung der Arbeit. Sie orientiert sich am Umfang dessen, was für eine einzelne Aufgabe zu leisten ist. Sie orientiert sich nicht an der Schwierigkeit der Aufgabe, sondern an der Anzahl der Schritte. Die schwerere Aufgabe bekommt nicht notwendigerweise mehr Punkte, sie bekommt in etwa gleich viele Punkte, wenn gleich viele Schritte zu erledigen sind; denn schwierigere Aufgaben, die von weniger Schülern gelöst werden, kommen bei der Umsetzung von Punkten zu Notenstufen sowieso zum Tragen.“ (GY, M)
- (9) „Bei Rechenfehlern, wenn der Ansatz richtig ist, gibt es dann Abzüge.“ (RS, M)
- (10) „Es hängt vom Schwierigkeitsgrad und der Länge der Aufgabe ab, ob entweder nur das Endergebnis gepunktet wird oder ob auch Rechenschritte gepunktet werden. Beides kommt vor.“ (FÖ, M)
- (11) „Je nach Schwierigkeit gibt es einen, zwei oder bei Textaufgaben auch drei oder mehr Punkte.“ (HS, M)
- (12) „Die Aufgaben mit hoher Punktzahl sind auch die schwierigeren.“ (HS, N)

B) „*Subtraktives Punkte-Modell*“ bedeutet, daß für einen bestimmten Aspekt z. B. zehn Punkte vorgegeben werden. Verstöße gegenüber festgelegten Anforderungen, Normen oder Erwartungen führen dann zu einem Punktabzug. In dieser Hinsicht korrespondiert das subtraktive Punkte-Modell der subtraktiven Form bei den Richtlinien-Modellen.

- (13) „Der Bewertungsmaßstab wird zumeist im groben Rahmen bei der Planung festgelegt, bei der Auswertung können dann Korrekturen erfolgen. D. h., wir legen Punkte fest und versuchen, nach diesen Punkten die Arbeit auszuwerten... Bei einer Nacherzählung sieht das so aus, daß wir einen Text haben... und daß wir festlegen, soviel Inhaltspunkte sind in diesem Text drin, das schreiben wir uns auch auf und bepunktet die Inhalte. Für die sprachliche Gestaltung gibt es soviel Punkte, wie müßte das aussehen? Wir besprechen das und legen das ziemlich genau in der Koordinationskonferenz fest... Das ist zunächst nur eine grobere Festlegung, daß wir z. B. sagen: bei dieser Nacherzählung sind zehn Punkte angemessen für Sprache und Ausdruck. Wenn jemand vom Satzbau her in ganzen, vollständigen Sätzen, mit Haupt- und Nebensatz schreibt, wenn keine häufigen Wiederholungen da sind, dann können wir sagen, das wären neun oder zehn Punkte. Dann acht oder sieben usw.“ (FÖ, D)

2.1.2.2. Fehler-Modelle

Die Fehler-Modelle sind dadurch gekennzeichnet, daß im Beurteilungsprozeß Fehler ausgezählt werden. Es lassen sich zwei Formen unterscheiden: Das *einfache-Fehler-Modell*, bei dem die gezählten Fehler bereits die Basis der Notengebung bilden, und das *Index-Fehler-Modell*, bei dem die Fehlerzahl im Hinblick auf eine andere gezählte Größe zu einem Indexwert (z. B. „Fehlerquotient“ als Zahl der Wörter insgesamt, dividiert durch die Zahl der Fehler) relativiert wird. Bei beiden Formen ist zusätzlich zu entscheiden, ob alle Fehler gleich gewichtet werden oder ob bestimmte Fehler höher (z. B. lernzielrelevante Fehler) oder niedriger (z. B. Interpunktionsfehler) gewichtet werden.

- (14) „Wir haben beschlossen, daß bestimmte Arten von Fehlern als 1,5 Fehler gewertet werden sollen... Ebenso haben wir definiert, was als leichter Fehler gelten soll und was für Wiederholungsfehler gilt.“ (GY, E)

Punkte- und Fehler-Modelle führen im Gegensatz zu den Richtlinien-Modellen nicht unmittelbar zu Noten, sondern es muß entschieden werden, welchen Punkt- bzw. Fehlerintervallen welche Notenstufen zugeordnet werden.

Im Zusammenhang mit dieser Zuordnung tauchen zusätzlich folgende Fragen auf:

- a) Orientiert sich die Zuordnung an den konkreten Ergebnissen der Klassenarbeit so, daß diese – z. B. in Form der Verwendung einer Normalverteilung für die Noten – die Zuordnungsvorschrift abgeben? (15)
 - b) Wenn die Zuordnungsvorschrift unabhängig vom konkreten Ausfall der Klassenarbeit festgelegt wird, erfolgt sie dann individuell durch den einzelnen Lehrer, oder liegen klassen- und/oder schulübergreifende Richtlinien vor? ((17), (18) und (15), (16))
 - c) Spielen wichtige Noten wie „sehr gut“ oder „ausreichend“ bei dieser Zuordnung eine besondere Rolle? ((16) bis (18))
- (15) „Den Bewertungsmaßstab legen wir teilweise vor der Arbeit fest, d. h., dann steht schon direkt auf dem Arbeitsblatt drauf: beim Erreichen von soviel Punkten ist das die Note ‚eins‘ usw. Bei anderen Arbeiten ist es so, daß den Schülern nur die Punktzahl vorgegeben ist für die einzelnen Aufgaben und daß wir dann, nachdem alle Kollegen die Arbeit nachgesehen haben, uns zusammensetzen und in Form einer Normalverteilung die Notengebung machen.“ (FÖ, E)
- (16) „Wir haben es in der Mathematik einfach, jeder Aufgabe eine bestimmte Punktzahl zuzuordnen, und können dann der Punktzahl wieder die Note zuordnen. Wir gehen in der Regel so vor, daß wir bei der Hälfte der zu erreichenden Punktzahl die Note ausreichend erteilen, gehen aber auch nach oben und nach unten, wenn es erforderlich ist, z. B. wenn wir erkennen: die Arbeit war wirklich zu schwierig oder sie war zu leicht, daß wir dann den Maßstab versetzen.“ (FÖ, M)
- (17) „Ich gehe etwa davon aus, daß 50% des Geforderten richtig sein müssen, um eine ausreichende Leistung zu erreichen. ... Ich gehe ja normalerweise von Fehlern aus. Es ist relativ selten, daß

ich Intervallschritte von drei bis vier Fehlern bei einer ‚Eins‘ habe; eventuell einen oder zwei Fehler, je nachdem, meistens sogar null Fehler.“ (GY, E)

- (18) „Ich habe im Grunde eine konstante Verteilung, die den Schülern auch bekannt ist. Von der Gesamtpunktzahl die ersten zehn Prozent sind eine ‚Eins‘, und der Rest wird dann aufgeteilt. Die Hälfte der zur unteren Grenze der ‚Eins‘ gehörenden Punkte kennzeichnen dann den Übergang zwischen ‚Vier‘ und ‚Fünf‘. Angenommen, wir haben 25 Punkte, dann wären die ersten drei Punkte, die etwa 10% entsprechen, eine ‚Eins‘. Dann 22, die Hälfte davon ist 11, 11 ist die Grenze zwischen ‚Vier‘ und ‚Fünf‘. Der Zwischenraum zwischen 22 und 11 wird dann in etwa gleiche Felder eingeteilt, der ‚Fünf‘ und der ‚Sechs‘ entsprechen ebenfalls gleiche Punktzahlen. Das ist so ein Grundraster, das ich immer anwende, und die Schüler stellen sich mittlerweile darauf ein.“ (GY, M)

Wenn bei der Zuordnung von Punkten- bzw. Fehlerintervallen zu Noten auch keine explizite Orientierung am Klassendurchschnitt wie bei der Anwendung einer Normalverteilung erfolgt, so stellt sich dennoch die Frage, ob die vorgegebenen Entscheidungen durch das konkrete Ergebnis der Klassenarbeit modifiziert werden können. Solche Modifizierungen können unterschiedliche Ursachen haben: So kann eine Arbeit insgesamt zu gut oder zu schlecht ausgefallen sein. Eine Über- oder Unterforderung mag erst bei der Korrektur sichtbar werden (16). Es kann aber auch so sein, daß sich an einer Intervallgrenze Punkte- bzw. Fehlerzahlen häufen, so daß etwa gleich leistungsstarke Schüler unterschiedliche Noten erhalten würden. Leichte Modifizierungen führen dagegen zu gleichen Noten bei etwa gleichen Leistungen (vgl. Zitate 19 bis 21). Und schließlich können solche nachträglichen Modifizierungen auch abgelehnt werden (22).

- (19) „Wenn ich sehe, daß beim Übergang von der ‚Vier‘ zur ‚Fünf‘, der z. B. bei zehn Punkten liegt, sehe, daß bei neun Punkten ein großer Teil liegt, dann nehme ich diese neun Punkte noch zur ‚Vier‘, mache aber eine ‚Vier minus‘ daraus.“ (FÖ, N)
- (20) „Wenn einem Schüler ein halber Punkt fehlt, dann ziehe ich lieber die Grenze einen halben Punkt runter, dann gebe ich lieber jemandem eine schlechte ‚Drei‘ als eine gute ‚Vier‘, soweit das möglich ist; wenn man in der Punktaufteilung Lücken hat.“ (FÖ, M)
- (21) „Die Hefte, die an den Grenzen der Intervalle liegen, die nehme ich mir immer noch einmal extra vor.“ (GY, M)
- (22) „Über die Punktvergabe mache ich mir schon bei der Planung Gedanken. Aber den Bewertungsmaßstab selbst, das mache ich unmittelbar vor der Korrektur. Da lege ich einen Zettel an – Punkte, Gesamtzahl – und dann wird die Einteilung nach Noten getroffen. . . Bevor ich das erste Heft in die Hand nehme, liegen die Noten fest, ohne Rücksicht darauf, wie die Arbeit ausfällt.“ (HS, M)

Unter diesem Aspekt ist auch die Frage bedeutsam, ob die endgültige Benotung unmittelbar nach der Korrektur dieses Heftes – also ohne Kenntnis der Punkte- oder Fehlerverteilung in der ganzen Klasse – erfolgt (23) oder ob erst nach Kenntnis dieser Verteilungen, also am Abschluß des Korrekturprozesses (vgl. Zitate 24 und 25).

- (23) „Wenn ich das Heft durchhabe, kommt sofort die Note darunter, denn die liegt ja fest.“ (HS, M)
- (24) „Die Note wird erst am Ende daruntergesetzt, wenn alle Arbeiten der Klasse korrigiert worden sind.“ (RS, M)
- (25) „Die Festlegung der Endnote erfolgt immer erst dann, wenn alle Arbeiten durchgesehen sind. Der Überblick muß erst da sein, obwohl dann schon die Voraussetzungen dafür festgehalten werden können: also Inhalt, Ausdruck usw. . . . Es müssen dann ja hinterher noch Vergleiche in horizontaler Richtung gemacht werden . . . und das ist erst dann sinnvoll, wenn die Arbeit fertig durchgesehen ist.“ (GY, E)

Im Zusammenhang mit der Festsetzung der Endnote ist noch zu fragen, ob die Note auch durch Faktoren, wie z. B. Rechtschreibung und äußere Form, beeinflusst wird.

- (26) „Die Rechtschreibung kann sogar bis zu einer Note Unterschied ausmachen. Wenn ich das Gefühl habe, der ist sonst, ich weiß das, der ist sonst ein relativ guter Rechtschreiber, und er gibt sich da überhaupt keine Mühe, dann kann es auch schon einmal passieren, daß ich ihm keine ‚Zwei‘, sondern nur eine ‚Drei‘ gebe und ihm sage: auch wenn wir einen Aufsatz schreiben, solltest du doch ein bißchen überlegen und darauf achten.“ (HS, D)
- (27) „Wenn einer sehr geschmiert hat, kriegt er auch schon einmal einen Punkt abgezogen.“ (RS, M)
- (28) „Rechtschreibfehler streiche ich zwar an – das muß ich –, aber die spielen bei der Bewertung keine Rolle.“ (GY, N)

2.2. Wiedergabe und Diskussion der Ergebnisse

Wiedergabe und Diskussion der Ergebnisse orientieren sich an dem in Abschnitt 2.1. beschriebenen Gerüst.

Bei der groben Unterscheidung nach Richtlinien- und Maßstabmodellen zeigen sich zunächst folgende Sachverhalte:

- Richtlinien-Modelle tauchen, abgesehen von einer Ausnahme (in der gymnasialen Mittelstufe werden ab und zu kleinere mathematische Betrachtungen geschrieben), nur in den sprachlichen Fächern auf; Mathematik und Naturwissenschaften sind durch Maßstab-Modelle gekennzeichnet.
- In den sprachlichen Fächern werden Richtlinien-Modelle nur bei bestimmten Arbeitsformen und/oder bestimmten Schulformen/Klassenstufen genannt. Im Fach Deutsch handelt es sich um den Deutschaufsatz (11/12 D)¹², im Fach Englisch werden vor allem in den höheren Klassenstufen der Realschule und des Gymnasiums Richtlinien-Modelle angesprochen (5/12 E), nicht aber in der Förderstufe und der Hauptschule.
- Im Fach Deutsch zeigt sich, daß bei Aufsätzen die inhaltliche Komponente eher in einer additiven Form (6/12 D) und die sprachliche Komponente eher in einer subtraktiven Form (8/12) beurteilt werden.
- Zwischen den Fächern Englisch und Deutsch scheint ein – hier nicht weiter belegbarer – Unterschied darin zu bestehen, daß in Englisch die Richtlinien – z. B. in Form von Fachkonferenzbeschlüssen – weit detaillierter ausgearbeitet sind als in Deutsch und zu einer größeren Beurteilungssicherheit bei sprachlichen Leistungen führen.
- Das Verhalten zwischen inhaltlicher und sprachlicher Komponente wird in Deutsch eher zugunsten der inhaltlichen Komponente entschieden, in Englisch eher zugunsten der sprachlichen.
- Die sprachlichen Fächer sind dadurch gekennzeichnet, daß in ihnen mehrere Beurteilungsmodelle verwendet werden.
Deutsch: Zusätzlich zum Richtlinien-Modell tritt bei den Diktaten das einfache Fehler-Modell und bei Grammatik/Kontrollarbeiten das additive Punkte-Modell auf.
Englisch: Hier taucht allein oder zusätzlich zum Richtlinien-Modell das additive Punkte-Modell auf (10/12 E). Fehler-Modelle werden selten allein, sondern überwiegend – in Form von Index-Fehler-Modellen – im Zusammenhang mit Richtlinien-Modellen diskutiert. In Mathematik und Naturwissenschaften taucht – abgesehen von der schon erwähnten Ausnahme – nur das additive Punkte-Modell auf.
- Das Fach Englisch nimmt hinsichtlich der Beurteilungsmodelle eine Mittelstellung zwischen Deutsch einerseits und den naturwissenschaftlichen Fächern andererseits ein.

Weit mehr und detailliertere Informationen liegen im Hinblick auf die Maßstab-Modelle vor: 41 der interviewten Lehrer informieren über Punkte-Modelle und zwölf über Fehler-Modelle, wobei von einigen Lehrern beide Formen angesprochen werden.

¹² Lies „11 von 12 interviewten Deutschlehrern“!

Es sollen zunächst die *Punkte-Modelle* betrachtet werden. Von subtraktiven Punkte-Modellen wird nur in zwei Fällen im Zusammenhang mit dem Deutschaufsatz berichtet (s. auch Zitat 13), wobei es sich bei einem dieser Fälle um ein früher praktiziertes Verfahren handelt, das jetzt nur noch ab und zu für Kontrollen eher pauschaler Entscheidungen verwendet wird. Bei den in der Schulpraxis auftretenden Punkte-Modellen handelt es sich somit fast ausschließlich um additive Punkte-Modelle, die im folgenden ausführlicher betrachtet werden sollen.

- a) 15 Lehrer (7/14 M; 8/13 N) geben an, daß sie versuchen, Aufgaben in Teilleistungen zu zerlegen, sei es in der Weise, daß die erforderlichen Teilschritte festgelegt werden, sei es, daß globaler von „Ansatz, Ausrechnung und Ergebnis“ gesprochen wird.
- b) Nicht untypisch ist der in Englisch, Mathematik und den Naturwissenschaften etwa gleich häufig angesprochene Fall, daß eine Aufgabe mit einer bestimmten Punktzahl belegt wird und Mängel oder Fehler zu Punktabzügen führen (4/12E; 4/12 M; 4/13 N). Es handelt sich hierbei um eine subtraktive Komponente im additiven Punkte-Modell.
- c) Im Regelfall wird für schwerere Aufgaben eine höhere Punktezahl vergeben; weit seltener wird der Fall genannt, daß die Anzahl der zu einer Aufgabe gehörenden Punkte unabhängig von der Schwierigkeit einer Aufgabe festgelegt werde und nur von der Anzahl der erforderlichen Schritte abhängen. Gleiche Punktzahlen werden vor allem in Englisch genannt; die Ursache dürfte in einer relativ hohen Aufgabenhomogenität liegen.

Bei der Frage der Zuordnung von Punktintervallen zu Notenstufen ist zunächst festzuhalten, daß diese nur in seltenen Fällen auf der Basis der konkreten Klassenarbeitsergebnisse – also z. B. am Durchschnittsergebnis orientiert, normalverteilt o. ä. – erfolgt. Nur in vier Fällen wurde von solchen – ausnahmsweise gewählten – Ansätzen berichtet.

Meistens wird die mögliche Gesamtpunktzahl nach jeweils relativ konstanten Kriterien und ohne Berücksichtigung des Ausfalls der Klassenarbeit vor Beginn der Korrektur in Punktintervalle zerlegt, denen dann die Noten zugeordnet werden. Im einzelnen zeigen sich folgende Ergebnisse:

- a) Es gilt das offensichtlich „eherne Gesetz“, daß für die Note „ausreichend“ etwa die Hälfte (50%) der möglichen Gesamtpunktzahl erforderlich ist (31/41, die mit additiven Punkte-Modellen arbeiten).
- b) Ein zweites, etwas schwächer ausgeprägtes „Gesetz“ scheint darin zu bestehen, daß zur Note „eins“ ein kleineres Punktintervall gehört als zu den übrigen positiven („zwei“ bis „vier“) Noten (26/41), wobei die prozentuale Größe dieses Intervalls variiert. In zwei Fällen liegen jedoch explizite Gegenäußerungen vor, hier ist das zur „Eins“ gehörende Punktintervall gleich groß wie bei den anderen positiven Noten.
- c) Läßt man die Note „eins“ außer Betracht, so variiert die prozentuale Breite der Punktintervalle zwar deutlich, aber die Mehrheit der Lehrer, die sich hierzu äußern (16/41), versucht, die Punktintervalle für die positiven Noten (außer „eins“) gleich breit zu machen; gleiches gilt für die negativen Noten „fünf“ und „sechs“.
- d) Die Zuordnung von Punktintervallen zu Noten ist vor allem in der Förderstufe und in Mathematik und den Naturwissenschaften im Gymnasium durch schulische Konventionen (z. B. in Form von Fachkonferenzbeschlüssen) geregelt (11/41). In zwei Fällen wird außerdem berichtet, daß der für die Sekundarstufe II geltende KMK-Schlüssel auch in der Sekundarstufe I angewendet werde.

Die vor dem Beginn des Korrekturprozesses vorgesehene Zuordnung von Punktintervallen zu Noten ist bei einem größeren Teil der Lehrer in Grenzen modifizierbar.

- a) 16 der 41 mit additiven Punkte-Modellen arbeitenden Lehrer geben an, daß die *reale* Punktverteilung (s. auch Zitat 20) zu leichten Modifizierungen des Schemas führen kann. Nur in vier Fällen wird dieses explizit abgelehnt, wobei die Ablehnung zum Teil mit der Lehrzielorientierung der Klassenarbeit begründet wird.

- b) Vier Lehrer berichten, daß bei der endgültigen Festlegung der Noten die 25%- bzw. die 50%-Klausel eine Rolle spiele. (Laut Erlaß müssen Arbeiten mit über 50% nicht ausreichenden Noten wiederholt werden; liegt der Anteil nicht ausreichender Arbeiten zwischen 25% und 50%, muß die Arbeit in einem schulinternen Verfahren genehmigt werden.) Aber zwölf Lehrer lehnen es explizit ab, aus diesem Grunde die Bewertung zu verändern.
- c) Sieben Lehrer berichten, daß sie bei erwartungswidrigem Ausfall den Schnitt leicht heben oder senken oder daß sie Teile, die von allen oder fast allen Schülern nicht bearbeitet wurden, aus der Wertung herausnehmen.

Mit den hier geschilderten Ergebnissen korrespondiert der Sachverhalt, daß 42 Lehrer die Benotung erst zum Abschluß der Korrektur durchführen; nur sechs Lehrer benoten unmittelbar nach Durchsicht eines Heftes. – Soweit die Ergebnisse im Hinblick auf die additiven Punkte-Modelle.

Bei den *Fehler-Modellen* ist noch die Frage der Fehlergewichtung offen, die nur in den Fächern Deutsch und Englisch, und zwar in zwölf Fällen, angesprochen wurde. Im Regelfall (Förderstufe und Gymnasium) werden Fehler gewichtet, wobei grundlegende Fehler oder lernzielrelevante Fehler ein höheres Gewicht erhalten (9/12). Gleiche Gewichtung wird nur von Lehrern der Hauptschulen angesprochen.

Bleibt als letzte Frage, die sich auf alle Beurteilungsmodelle bezieht, wie sich Rechtschreibung und äußere Form auf die Gesamtnote auswirken. Die äußere Form wird nur dreimal in Mathematik thematisiert, dort kann eine schlechte äußere Form zu Punktabzügen führen. Differenzierter ist das Bild bei der Rechtschreibung. Rechtschreibfehler führen bei Englisch-Arbeiten, auch wenn es von der Sache her nicht primär um Rechtschreibung geht, zu Punktabzügen (6/12 E). Im Deutschaufsatz werden Rechtschreibfehler von Lehrern korrigiert (8/12 D), sie spielen aber bei der Beurteilung meist nur eine eingeschränkte Rolle (6/12 D); in zwei Fällen wird explizit ein Eingehen der Rechtschreibleistung in die Aufsatznote abgelehnt. In Mathematik und in den Naturwissenschaften scheint die Rechtschreibung von geringer Bedeutung: nur zwei Lehrer berichten, daß sie – aus Pflichtgefühl – die Rechtschreibfehler korrigieren, die Rechtschreibfehler aber für die Bewertung irrelevant seien.

Welches Fazit läßt sich nun aus diesen Ergebnissen ziehen? Zusätzlich zu der in Abschnitt 2.1. versuchten Systematisierung werden insbesondere die folgenden Sachverhalte als wichtig angesehen:

Es zeigen sich deutlich einige fach- und/oder schulformenspezifische Tendenzen. Unter der hier geteilten Voraussetzung, daß sich diese Lehrerverhaltensweisen in der Praxis bewährt haben, kann eine wichtige und lohnende Zielsetzung von Ausbildungsprozessen darin bestehen, diese bewährten Konzepte theoretisch und praktisch aufzuarbeiten.

Bei differenzierterer Betrachtungsweise läßt sich aber auch eine große Variabilität zwischen den einzelnen Lehrern, selbst bei gleichem Fach und/oder gleicher Schulform, aufzeigen. Zwar verhalten sich die interviewten Lehrer – was an dieser Stelle nicht weiter belegt werden kann – meist äußerst konsistent in der Handhabung ihres Bewertungsansatzes: wenn, was hier angenommen wird, Interviewverhalten und Realverhalten nicht wesentlich auseinanderfallen, sind sie jeweils für den Schüler einschätzbar. Aber es wird dennoch deutlich, daß Lehrer individuell unterschiedliche Beurteilungsstile entwickeln; gleiche fach- und/oder schulformenspezifische Fragen werden auf individuell unterschiedliche Weise beantwortet.

Hier zeigt sich auf der Basis der schulischen Beurteilungspraxis, warum die Notengebung bei genereller Betrachtung unzuverlässig und wenig gültig erfolgt. Indessen: eine generelle Erhöhung von Reliabilität und Validität hätte eine Ausschaltung der individuellen Komponente und damit eine Normierung von Unterrichts- und Beurteilungsprozessen zur Voraussetzung – eine wenig schöne Vision!

Positiv gewendet heißt dies, von der Forderung, Noten müßten generell vergleichbar sein, Abschied zu nehmen. In einem Interview wurde diese Konsequenz bündig formuliert:

- (29) „Ich sage mitunter den Schülern: ‚Im Grunde genommen, die Note ‚drei‘, die ich im Zeugnis erteile, ist eine ‚drei‘ bezogen auf den Stoff, bezogen auf die Schulf orm, bezogen auf die Altersstufe und bezogen auf mich, auf den Lehrer.‘ Und das läßt sich nicht ohne weiteres – leider – mit anderen wieder vergleichen. Da gibt es mit Sicherheit Schwankungen.“ (RS, D)

Die Forderung an Lehrer, *ein und dieselbe* Schülerleistung unabhängig von ihren Entstehungsbedingungen, losgelöst von einem spezifischen Unterrichts-, Klassen- und Lehrerkontext mit *gleicher* Note zu beurteilen, erweist sich damit als wenig sinnvoll. (Dieser Sachverhalt macht deutlich, warum solche Untersuchungen zur Reliabilität der Notengebung wenig hilfreich sind, die die Möglichkeit gleicher Beurteilungen ohne Kenntnis der Entstehungsbedingungen der Schülerleistung voraussetzen.) Lehrer können nur die Forderung erfüllen, daß ihre Beurteilung *unterrichtsreliabel* und *unterrichtsvalide* bzw. allgemeiner *kontextreliabel* und *kontextvalide* sind. Wird dieser Anspruch akzeptiert, so hat dies sicher eine für den Lehrer sehr entlastende Funktion.

Diese Einschätzungen dürfen nicht in der Weise mißverstanden werden, daß dann Leistungsbeurteilung überflüssig sei oder daß es dem individuellen Belieben eines Lehrers überlassen sei, was er wie bewerte. Bezogen auf den Kontext, in dem Schülerleistungen in Form von Klassenarbeiten erbracht werden, ist es wichtig zu wissen, ob die Schüler den hier gestellten Anforderungen genügen und ob – diese Kehrseite der Medaille wird häufig vergessen – die Lehrer einen angemessenen Unterricht erteilt haben. Außerdem existieren unabhängig von der individuellen Lehrperson Lehrpläne, Rahmenrichtlinien, Konferenzen, Schulbücher usw., deren normative Kraft als ausreichend dafür angesehen werden kann, daß Lehrer nicht Beliebiges in beliebiger Weise beurteilen.

2.3. Zur Beurteilungssituation im Fach Deutsch: Aufsatzbeurteilung

In diesem Abschnitt sollen auf der Basis von Interviewaussagen Probleme der Beurteilung von Deutschaufsätzen gesondert angesprochen werden. Zwar werden von den Anwendern additiver Punkte-Modelle durchaus Probleme, die mit deren Anwendung verknüpft sind, gesehen:

- (30) „Dieses Belegen mit Punkten ist ja auch ganz sicher eine Sache, die manipulierbar ist; gerade im Hinblick auf diese 50%-Klausel – bei 50% muß die Arbeit wiederholt werden, zwischen 25% und 50% muß sie genehmigt werden – da läßt sich sicher mit der Zuordnung manipulieren. Auch die Bewertung der einzelnen Aufgaben läßt sich hochspielen oder runterjubeln usw., so daß man sicher sagen kann: das Belegen der einzelnen Aufgaben mit Bewertungseinheiten ist eine Hilfe für die Korrektur, aber im Grunde liefert sie nur eine Scheinobjektivität.“ (GY, M)

Gleichwohl läßt sich sagen, daß die Verwendung additiver Punkte-Modelle und die damit oft verknüpfte Reduktion der Beurteilungsentscheidungen auf einfache „Richtig-Falsch-

Entscheidungen“ oder „Vorhanden-Nicht vorhanden-Entscheidungen“ den Korrekturprozeß für den Lehrer vereinfachen und für den Schüler durchschaubar machen. Sieht man von Ausnahmefällen wie z. B. der „pädagogischen Note“ einmal ab, so kommt hinzu, daß alle Schüler am gleichen „Maßstab“ gemessen werden. Ergebnisse dieses Prozesses sind bei Kenntnis des Maßstabes leicht intersubjektiv überprüfbar.

Diese Vorteile sind im Rahmen von Richtlinien-Modellen zunächst nicht gewährleistet. Es gibt Sachverhalte, die sich nicht in Form von Punkten quantifizieren lassen, die aber dennoch nach geltender Erlaßlage benotet werden müssen.

- (31) „Ich sehe auch keine Ziffern, wenn ich die Arbeit lese, sondern ich frage mich: ‚Wie schätzt du was ein?‘, und sage dann: ‚Nach dem, was wir gemacht haben, ist das, was der dort geschrieben hat, wirklich gut.‘ Und wenn Mängel da sind, schlimme Mängel, dann muß ich eine Note heruntergehen... Es gibt keine Punktbewertung bei solchen Aufsätzen. Sie können den Gesamteindruck eines solchen Aufsatzes nicht losgelöst sehen von dem, was vorher war.“ (HS, D)

Der Versuch, hier mit subtraktiven Punkte-Modellen (s. Zitat 13) zu arbeiten, ist eher die Ausnahme als die Regel. Es kommt hinzu, daß hier, wenn auch in anderer Form, eine im Prinzip gleiche Situation vorliegt: wann genau ziehe ich einen Punkt ab, wann zwei usw.?

Welche Konsequenzen ziehen Deutschlehrer aus der insbesondere bei der Beurteilung von Deutschaufsätzen schwierigen Situation? Drei solcher Konsequenzen lassen sich in den Interviews nachweisen:

- a) Momente von Unsicherheit/Unzufriedenheit mit den beruflichen Anforderungen.

- (32) „Ich kann natürlich sagen, dieser Aufsatz gefällt mir besser als der. Aber dann traue ich mich überhaupt keine Note mehr zu geben. Wer legitimiert mich dazu, daß ich darüber urteilen soll? Auch deshalb fallen mir die Aufsätze so schwer.“ (FÖ, D)

- (33) „In meinem zweiten Leben werde ich bestimmt kein Deutschlehrer mehr. Das ist sehr, sehr schwierig.“ (FÖ, D)

- b) Unsicherheit in der Notengebung, Vermeidung von Extremnoten, Konzentration auf den mittleren Notenbereich.

- (34) „Bei Aufsätzen habe ich, wenn ich mich selber überprüfe, eine Tendenz zur Mitte. Ich habe meine Notenspiegel angesehen, es zeigt sich, daß die Noten ‚drei‘ und ‚vier‘ fast 70% ausmachen.“ (RS, D)

- (35) „Im Aufsatz ist es schwierig, weil, und davon gehe ich aus, wenn jemandem gezeigt wird, du bist im Deutschen, also deiner Sprache, ungenügend, dann habe ich ein bißchen Sorge, daß man den Schüler sehr trifft. Es gibt einige Arbeiten, die eindeutig ‚fünf‘ sind, aber mehr als eine Fünf gebe ich eigentlich kaum. Das wird dann eher so um die Vier herum gruppiert.“ (GY, D)

- (36) „Es wird dann so sein, daß beim Aufsatz selten eine Sechs herauskommt, das werden dann immer meistens die Noten sein zwischen ‚zwei‘ und ‚fünf‘ – schlechtestenfalls.“ (HS, D)

- c) Bevorzugung solcher Themen- und Aufgabenstellungen, die als leichter beurteilbar gelten.

- (37) „Da gibt es ja deutliche Kriterien bei unterschiedlichen Aufsatzformen. Z. B. bei der Inhaltsangabe ist es relativ klar: da gibt es wichtige Elemente, die auf jeden Fall genannt werden müssen... Der Schüler weiß auch, daß er unwichtige Teile nicht schreiben sollte, das wäre dann eine Nacherzählung... Da notiere ich mir dann die Gesichtspunkte, und das kann man auch abhaken. Deswegen sind Inhaltsangaben auch so beliebt, weil man da etwas ganz deutlich Konkretes hat, und das leuchtet auch jedem Schüler ein.

Schwieriger wird es mit der Bewertung der Ausdrucksfähigkeiten. Da kann man nicht mehr zählen, da geht es um den Eindruck, den man von dem Text hat.“ (GY, D)

- (38) „Da hatte ich mir, bevor ich ans Korrigieren ging, gedacht: das muß drin sein. Diese Informationen enthielt der Text, und die mußten die Schüler dem Text entnehmen und selbst verwerten. Dann konnte ich sagen: da ist es gut und da nicht so – abgesehen vom sprachlichen Aspekt. Deswegen ist mir so etwas lieber als nur so reine Gefühlsaussagen.“ (FÖ, D)

Diese Aussagen sind für die interviewten Deutschlehrer vermutlich nicht repräsentativ, sie belegen aber mindestens symptomatisch die eingangs genannte schwierige Situation.

Ist damit eine Beurteilung, die von allen Beteiligten intersubjektiv nachvollzogen werden kann, bei Deutschaufsätzen nicht mehr möglich? Eine große Zahl von Hinweisen in den Interviews zeigt, daß solche Möglichkeiten durchaus bestehen. Hier verläuft die Beurteilung eher in Form eines Vorgangs, der sich vorwiegend am Lehrerschreibtisch abspielt. In diesem Prozeß werden die Beurteilungskriterien nicht, wie häufig in der Mathematik und den Naturwissenschaften, von außen an das zu beurteilende Produkt herangetragen, sondern hier stellt die Erörterung der Beurteilungskriterien eine unterschiedliche Aufgabe dar. Es kommt hinzu, daß der Korrekturvorgang selbst so abläuft, daß er in Form eines Konvergenz-Prozesses beschrieben werden kann.

Ein Beurteilungsprozeß der geschilderten Art setzt zunächst voraus, daß die Beurteilungskriterien Gegenstand des Unterrichts sind, wobei jene Form, bei der die Kriterien im Unterricht erarbeitet und dort von den Schülern auf ihre eigenen Arbeitsprodukte angewandt werden, vermutlich die pädagogisch sinnvollste und zugleich effektivste sein dürfte.

- (39) „Bei einer Bildergeschichte z.B. werden auch die Punkte, die Kriterien, vorher genau festgelegt, die erfüllt sein müssen. Und diese Kriterien werden dann auch mit den Kindern während der Einheit besprochen. Wenn wir also Bildgeschichten gemeinsam besprechen, listen wir auch diese Kriterien an der Tafel auf – sie werden sogar oft gemeinsam mit den Kindern erstellt. Wir versuchen dann, Arbeiten, die als Hausaufgaben oder in der Klasse entstanden sind, auch nach den Kriterien zu beurteilen. Das ist also ein bißchen transparent für Kinder, was da geschieht“ (FÖ, D)
- (40) „Bei den Aufsätzen machen wir es meist so, daß wir kurz vorher schon einmal Probeaufsätze von den Schülern bewerten lassen. Also die und die Sachen, die müßt ihr jetzt gut, mittel oder schlecht ankreuzen. Und sie erarbeiten erst die Bewertungskriterien, und dann gehen sie ihre eigenen Aufsätze meistens schon einmal durch und wissen, wie es ungefähr läuft.“ (FÖ, D)

Es ist offensichtlich, daß auf diese Weise die Kriterien für die Beteiligten durchschaubar und handhabbar werden.

An zweiter Stelle ist der Korrekturprozeß selbst zu nennen. Bei der Korrektur von Deutschaufsätzen verläuft dieser Prozeß mindestens in zwei, im Regelfall in drei, oft aber auch in noch mehr als drei Durchgängen durch alle Hefte einer Klassenarbeit. Jeder Durchgang erfüllt für den Lehrer eine spezifische Funktion: jedoch variieren diese Funktionen deutlich von Lehrer zu Lehrer. Sie sind aber fast immer so konzipiert, daß in ihren Vollzug die Entscheidungsfindung bezüglich der Note eingebaut ist.

- (41) „Ich nehme die Arbeiten und streiche zunächst nur alle Fehler an; beim ersten Durchgang... Dabei lese ich wohl schon, aber noch nicht sehr bewußt. Und im zweiten Durchgang, dann lese ich einfach nur, ohne etwas anzustreichen. Beim Lesen versuche ich, mir schon ein paar Häufchen zurechtzulegen. Dann habe ich so meistens drei Häufchen, eines mit den guten, eines mit den mittleren und eines mit den schlechten Arbeiten. Dann gehe ich jedes Häufchen noch einmal einzeln durch. Bei diesem direkten Vergleich entdecke ich oft, daß ich mich geirrt habe.“ (HS, D)

Zusammenfassend bewertet ein Lehrer ein solches Vorgehen als

- (42) „einen sich ständig ver- und absichernden Prozeß.“ (FÖ, D)

Den Abschluß des Beurteilungsprozesses bildet die Rückgabe der Klassenarbeit, bei der u. a. Kriterien- und Beurteilungsfragen selbst wieder in diskursiver Form Gegenstand des Unterrichts sind.

- (43) „Ich lese eine gute und eine weniger gute Arbeit vor, erläutere, warum die eine Arbeit gut und warum sie besser als die andere ist. Ich erkläre auch Fehler, die sich gehäuft haben.“ (RS, D)
- (44) „Ich lese zunächst drei Aufsätze – gut bis schlecht – im Unterricht vor, die werden im Unterricht zunächst gründlich besprochen. Und dann erst teile ich die Hefte aus.“ (FÖ, D)

Für Schüler, deren Leistungen in der hier idealtypisch geschilderten Form beurteilt werden, ist der Beurteilungsprozeß sicher auch transparent nachvollziehbar und – aus Schülersicht – von Willküraspekten befreit.

Es wäre eine reizvolle weitere Aufgabe zu überlegen, ob Momente dieses Prozesses, insbesondere im Hinblick auf die Vorbereitung einer Klassenarbeit im Unterricht und die Rückgabe der Klassenarbeiten, nicht auch auf andere Fächer – Mathematik, Naturwissenschaften – übertragbar sind und welche Konsequenzen sich aus einer solchen Veränderung für Unterrichtsgestaltung und Beurteilungsprozesse ergeben würden.

Anschrift des Autors:

Prof. Dr. Willi Wolf, Fachbereich Erziehungswissenschaften, Philipps-Universität Marburg, Wilhelm-Röpke-Straße 6, 3550 Marburg 1.